Functional Object Re-use and Exchange: Supporting Information Topology Experiments

http://foresite.cheshire3.org/

Rob Sanderson  (azaroth@liverpool.ac.uk)
Richard Jones    (richard.d.jones@hp.com)
Clare Llewellyn  (clare.llewellyn@jstor.org)
Outline

Introduction to Foresite
  What is the project about?
  How far have we got?
Foresite software libraries
  Architecture and Object Model
  Serialisations
  How to get involved
SWORD and ORE
DSpace and ORE
  SWORD
  ORE objects as DSpace items
  Identifiers and embedded RDFa
  An ORE interface
What is the project about?

Grand Vision:
Bootstrap ORE-based scholarly communication processes

Two phases:
Describe journal/issue/article hierarchy in ORE
Import descriptions into DSpace

Requirements:
Large collection of scholarly communication described with ORE
DSpace to understand & ingest ORE Resource Maps
DSpace to allow linking back to original source

JSTOR:
1000+ journals, 185,000+ issues, 1.8M+ articles
Described in XML, down to OCR of article text
Introduction to Foresite
Introduction to Foresite

Diagram: Cylinders represent JSTOR databases, with arrows pointing to related elements. Domains lead to journals, which lead to issues. Issues lead to articles, which have sections, experimental resources, data, words, pages, figures, and titles.
How far have we got?

Generate ORE Resource Maps from JSTOR data
Prototyping complete; initial serialisations generated; main export development underway

Construct ORE Libraries in Java and Python
initial versions (0.9) released; object model complete; serialisers/parsers nearing completion; Atom still in question

Modify DSpace and SWORD to ingest resource maps
SWORD correctly interprets passed resource maps

Modify DSpace to manage and expose ORE Resource Maps
Stores and transformes ingested resource maps; presents items based on Resource Maps; embeds RDFa in items; interfaces for other serialisations

Conduct data exchange experiments to demonstrate the standard and the project outputs pending
Outline

Introduction to Foresite
  What is the project about?
  How far have we got?
Foresite software libraries
  Architecture and Object Model
Serialisations
  How to get involved
SWORD and ORE
DSpace and ORE
  SWORD
  ORE objects as DSpace items
  Identifiers and embedded RDFa
  An ORE interface
Architecture and Object Model
Aggregation agg = OREFactory.createAggregation(uri_a);

ResourceMap rem = agg.createResourceMap(uri_r);

AggregatedResource ar = agg.createAggregatedResource(uri_ar);

Proxy proxy = ar.createProxy(uri_p);

Triple triple = agg.createTriple(predicate, uri_object);
Serialisations

ATOM: currently still under discussion
RDFa: microformat for HTML page embedding
RDF: various serialisations RDF/XML, N3, Turtle, N-Triples

ORESerialiser s = ORESerialiserFactory.getInstance("RDF/XML");
ResourceMapDocument rmd = s.serialise(resourceMap);

OREParser p = OREParserFactory.getInstance("N3");
ResourceMap = p.parse(inputStream);
Getting Involved

Website: http://foresite.cheshire3.org/

Wiki: http://foresite.cheshire3.org/wiki

Google Code: http://foresite-toolkit.googlecode.com/

Google Group: http://groups.google.com/group/foresite
Outline

Introduction to Foresite
  What is the project about?
  How far have we got?
Foresite software libraries
  Architecture and Object Model
  Serialisations
  How to get involved

**SWORD and ORE**

DSpace and ORE
  SWORD
  ORE objects as DSpace items
  Identifiers and embedded RDFa
  An ORE interface
SWORD: Simple Web-service Offering Repository Deposit

JISC funded between 1 March and 31 October 2007
Profile of Atom Publishing Protocol

Simple Case:
Repository publishes self-describing service document
Client POSTs data (with HTTP headers) to Repository
Repository responds with an <atom:entry> document

Less Simple Case:
Client POSTs data on behalf of user known to Repository
Repository authenticates and responds with (more complex) <atom:entry> document
3 methods for combing SWORD and ORE

1 – Pass URL of Resource Map as part of the SWORD header

2 – Pass serialised Resource Map as the SWORD deposit package

3 – Pass ZIP file using serialised Resource Map as package manifest
Outline

Introduction to Foresite
  What is the project about?
  How far have we got?
Foresite software libraries
  Architecture and Object Model
  Serialisations
  How to get involved
SWORD and ORE
DSpace and ORE
  SWORD
  ORE objects as DSpace items
  Identifiers and embedded RDFa
  An ORE interface
Using SWORD + ORE in DSpace

Add new Ingester to deal with ORE Resource Maps as packages attached to SWORD deposit

Store incoming Resource Map

Localise Resource Map using DSpace identifiers, and crosswalk to canonical form (RDF/XML)

Extract metadata from Resource Map and populate metadata record
ORE Objects as DSpace Items

Store localised Resource Map in ORE bundle

Intervene in Bundle READ process, read in localised Resource Map, and generate DSpace Item object based on its content

Send “fake” Item to UI for rendering as normal
Identifiers and embedded RDFa

**what does an identifier identify?**

Is it the digital object?  
Is it the splash page?  
Is it the Resource Map?

**How do we get the Resource Map?**

Understand that an object identifier and the object URL may be different  
Insert `<link>` tags in the splash page  
Turn the splash page INTO a Resource Map...
Identifiers and embedded RDFa

`<div id="ore:ResourceMap">
  <div about="this-page.html">
    <a rel="rdf:type"
       href="http://www.openarchives.org/ore/terms/ResourceMap"></a>
    <a rel="rdf:type" href="info:eu-repo/semantics/humanStartPage"></a>
    <a rel="ore:describes" href="myobject/aggregation"></a>
  </div>
  <div about="myobject/aggregation">
    <a rel="ore:isDescribedBy" href="myobject/rdfxml"></a>
    <a rel="ore:aggregates" href="blabla/resource1.pdf"></a>
  </div>
  <div about="myobject/resource1.pdf">
    <span property="dc:title" content="My Title"></span>
  </div>
</div>`
An ORE Interface

http://www.mydspace.ac.uk/ore/atom/hdl/123456789/100
http://www.mydspace.ac.uk/ore/ntriple/hdl/123456789/100
http://www.mydspace.ac.uk/ore/n3/hdl/123456789/100
http://www.mydspace.ac.uk/ore/rdfxml/hdl/123456789/100
http://www.mydspace.ac.uk/ore/turtle/hdl/123456789/100
Richard Jones
Hewlett-Packard Laboratories
richard.d.jones@hp.com